



Article Forward Greedy Searching to κ-Reduct Based on Granular Ball

Minhui Song, Jianjun Chen *, Jingjing Song, Taihua Xu and Yan Fan

School of Computer, Jiangsu University of Science and Technology, Zhenjiang 212100, China * Correspondence: jianjunchen@just.edu.cn

Abstract: As a key part of data preprocessing, namely attribute reduction, is effectively applied in the rough set field. The purpose of attribute reduction is to prevent too many attributes from affecting classifier operations and reduce the dimensionality of data space. Presently, in order to further improve the simplification performance of attribute reduction, numerous researchers have proposed a variety of methods. However, given the current findings, the challenges are: to reasonably compress the search space of candidate attributes; to fulfill multi-perspective evaluation; and to actualize attribute reduction based on guidance. In view of this, forward greedy searching to *k*-reduct based on granular ball is proposed, which has the following advantages: (1) forming symmetrical granular balls to actualize the grouping of the universe; (2) continuously merging small universes to provide guidance for subsequent calculations; and (3) combining supervised and unsupervised perspectives to enrich the viewpoint of attribute evaluation and better improve the capability of attribute reduction. Finally, based on three classifiers, 16 UCI datasets are used to compare our proposed method with six advanced algorithms about attribute reduction and an algorithm without applying any attribute reduction algorithms. The experimental results indicate that our method can not only ensure the result of reduction has considerable performance in the classification test, but also improve the stability of attribute reduction to a certain degree.

Keywords: approximation quality; attribute reduction; conditional entropy; granular ball; rough set; sample division



check for

Citation: Song, M.; Chen, J.; Song, J.; Xu, T.; Fan, Y. Forward Greedy Searching to κ-Reduct Based on Granular Ball. *Symmetry* **2023**, *15*, 996. https://doi.org/10.3390/sym15050996

Academic Editors: Florentin Smarandache, Dalibor Štys and Sergei D. Odintsov

Received: 23 March 2023 Revised: 22 April 2023 Accepted: 25 April 2023 Published: 27 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

With the rapid increase in data, attributes have become redundant and uncertain. Uncertainty mainly consists of the five following aspects: incompleteness, inconsistency, incompatibility, fuzziness and randomness. Consequently, finding valuable information from high-dimensional data is a challenge for the research field.

With the intention of effectively disposing of ambiguous, incomplete, and inaccurate data, Polish scholar Pawlak first put forward rough set theory [1] in 1982, which has been extensively adopted in data mining, pattern recognition, decision analysis [2–4], and other domains. Based on rough set theory, many extensions and improvements have been proposed, such as neighborhood rough set [5], fuzzy rough set [6], decision-theoretic rough set [7], and Pythagorean fuzzy set [8]. Attribute reduction [9–13], as a common dimensional reduction method, can efficaciously remove redundant components in information systems, choose the optimal minimum attribute subset and further improve the effectiveness of data knowledge discovery. Obviously, attribute reduction has grown to be a paramount research branch of rough set theory.

Generally speaking, simplified searching strategies can be split into two general classes: exhaustive search and heuristic search [14]. In the process of data analysis, the final reduction result is directly related to the given constraint which can be implemented by constructing different measurement criteria.

As a mature heuristics-based search, The forward greedy strategy has a wide range of measures [15–17], such as approximation quality and conditional entropy [18–21]. These measures are especially used for assessing attributes and exporting reduction results.

However, in the exploration of attribute reduction, many researchers only take into account the single view measure to determine the constraint. For instance, Jiang et al. [22] studied the supervised neighborhood attribute reduction; Zhang et al. [23] investigated a semi-supervised attribute reduction method that combined the collaborative learning theory; and Yuan et al. [24] introduced a fuzzy complementary entropy measure and proposed an unsupervised attribute reduction algorithm for mixed data. In order to fully consider the diversity of evaluation, multi-view measures are necessary to be proposed in attribute reduction.

The neighborhood rough set provides a flexible granular representation, but it requires determining the size of a neighborhood radius through grid search frequently, which is timeconsuming. To overcome this problem, many strategies for determining the radius without parameters have been introduced. As an example, Xia et al. [25] put forward the concept of granular balls and enhanced the efficacy of a classifier based on granular computing by generating granular balls; Zhou et al. [26] proposed the concept of a gap neighborhood when resolving the problem of online feature selection, which can automatically determine the neighborhood size according to the distance difference between samples.

Through the above discussion, in order to effectively obtain the salient features of multiple views [25–27] and improve the classification performance of attribute reduction, we propose a new strategy in this paper: forward greedy searching to κ -reduct based on granular ball. The key of our strategy includes three phases: (1) grouping the samples in the whole universe (the universe is a finite set of all samples) based on generated symmetric granular balls; (2) attribute search for sample groups based on guidance; and (3) attribute reduction based on multiple perspectives. The first stage is to automatically create granular balls in accordance with the distribution of data itself, and merge granular balls which have small number of samples. Therefore, it can realize the division of the universe. In addition, the second stage is guidance-based evaluations, aiming to compress the search space related to candidate attributes. Therefore, the time needed for attribute reduction can be made faster because fewer candidate attributes need to be evaluated. Finally, the third stage blends supervised and unsupervised perspectives [28] and uses the quality-to-entropy ratio as a measure to attribute reduction. Therefore, it is feasible to identify attributes and labels with accuracy, and quantitatively characterize the uncertainty of data itself [29,30].

To sum up, the main contributions of our research are: (1) decreasing the size of samples by grouping different granular balls which adaptively generate; (2) enhancing attribute reduction efficiencies by achieving guidance-based search; and (3) utilizing the quality-to-entropy ratio which combines two perspectives to improve the accuracy of recognizing eligible attributes.

The remainder of this paper is organized as follows. Section 2 introduces the basic concepts of rough set, granular computing, and attribute reduction. Section 3 describes the fundamental framework and specific procedures of the new proposed method. Comparative experimental results of datasets and analysis are reported in Section 4. Finally, Section 5 is a summary of the algorithm and points for further work.

2. Preliminaries

2.1. Neighborhood Rough Set

Formally, a decision system can be defined as a binary group, expressed as $DS = \langle U, AT \cup \{d\} \rangle$: the universe of discourse $U = \{x_1, x_2, \dots, x_n\}$ is a set of non-empty finite samples; $AT = \{a_1, a_2, \dots, a_m\}$ is the set of all conditional attributes; and *d* is the decision attribute. According to the decision values of all samples, it is not difficult to obtain a partition like $U/IND(d) = \{X_1, X_2, \dots, X_n\}$ which induced by decision attribute *d* on universe *U*: $IND(d) = \{(x_i, x_j) \in U \times U : d(x_i) = d(x_j)\}; \forall x_i \in U, d(x_i)$ is the label of sample x_i . It is especially worth noting that IND(d) is a relation of equivalence with symmetry, reflexivity, and transitivity. The following definitions are the form of conventional rough sets.

Definition 1. For a given decision system DS, a given radius $\delta \ge 0$, $\forall A \subseteq AT$, $\forall x_i \in U$, the neighborhood of $\delta_A(x_i)$ is defined as:

$$\delta_A(x_i) = \{ x \in U : \Delta_A(x_i, x_j) \le \delta \},\tag{1}$$

in which $\Delta_A(x_i, x_i)$ represents the distance function between sample x_i and x_i with respect to A.

Immediately, from Definition 1, it can be obviously known that the size of the generated neighborhood relies on the value of given δ , i.e., the neighborhood becomes larger as the value of δ increases.

As the fundamental units of the neighborhood rough set, the specific definitions of upper and lower approximations are given in the following Definition 2.

Definition 2. For a given decision system DS, $\forall X \subseteq U$, $\forall A \subseteq AT$, the lower and upper approximations of X are defined as

$$\underline{N}X = \{x_i \in U : \delta_A(x_i) \subseteq X\},\tag{2}$$

$$\overline{N}X = \{x_i \in U : \delta_A(x_i) \cap X \neq \emptyset\}.$$
(3)

The neighborhood rough set is built on the foundation of the standard rough set. It can not only deal with complex data, but also possess a multi-granularity structure by giving various radii. However, finding the appropriate radius generally requires a large number of trials or a certain parameter searching strategy, which is very time-consuming.

2.2. Granular Ball Computing

Considering that using the neighborhood relationship takes significant time to obtain the optimal radius, Xia et al. [31] proposed the concept of granular ball. In rough set theory, granule is the division of a sample set, and granular ball is based on the concept of granule. Xia et al. [25] regard hyper-ball with a completely symmetrical structure as granular ball.

The granular ball has a straightforward geometric shape with two parameters, i.e., center and radius. Compared with the neighborhood, the granular ball method has higher searching efficiency and robustness. The detailed definitions are as follows.

Definition 3. For a given decision system DS, $\forall A \subseteq AT$, $\forall GB_S \subseteq U$, GB_S is a granular ball induced by conditional attribute set A if and only if C is the center point of GB_S , r is the average of distances from all samples in the granular ball to C. The C and r of the granular ball are expressed as follows

$$C = \frac{1}{|GB_S|} \sum_{i=1}^{|GB_S|} x_i,$$
(4)

$$r = \frac{1}{|GB_S|} \sum_{i=1}^{|GB_S|} \Delta_A(x_i, C),$$
(5)

in which $|GB_S|$ indicates the number of samples in the granular ball.

In the following, GB_A^U is defined as the set of all granular balls induced by conditional attribute set *A* on universe *U*.

Definition 4. For a given decision system DS, $\forall A \subseteq AT$, $\forall GB_S \subseteq GB_A^U$, $d(GB_S)$ is recorded as the overall label of GB_S , i.e., $d(GB_S)$ is the label corresponding to samples with the same label and maximum proportion in the granular ball.

Definition 5. For a given decision system DS, $\forall A \subseteq AT$, $\forall GB_S \subseteq GB_A^U$, the average purity of GB_S is defined as

$$Pure(GB_s) = \frac{|\{x_i \in GB_s : d(x_i) = d(GB_s)\}|}{|GB_s|}.$$
(6)

in which $d(x_i)$ indicates the label of the sample x_i .

Furthermore, $Pure(GB_A^U)$ can be recorded as the mean purity of all granular balls induced by conditional attribute set *A*.

In the process of generating granular balls, the main idea is using an iterative twomeans algorithm. The concrete procedures are given as follows.

(1) Consider the entire universe *U* as an initial granular ball and set n = 1 (*n* is the number of existing granular balls).

(2) Cluster each cluster by the two-means algorithm.

(3) Compute the center point of each cluster and the average distance between each cluster's samples and the center point.

(4) Obtain the granular ball and calculate the granular ball's purity.

(5) Traverse all currently existent granular balls; if each granular ball's purity is below the given threshold, end this step; otherwise, return to (2).

On the basis of the aforementioned method of obtaining granular balls, Xia et al. [31] further put forward the concept of granular ball rough set, as shown in Definition 6.

Definition 6. For a given decision system DS, $\forall A \subseteq AT$, $\forall X_P \in U/IND(d)$, according to the conditional attribute set A, the upper and lower approximations of X_p are, respectively, defined as

$$\overline{GB_A^U}(X_p) = \Big\{ x_i : \exists GB_S \in GB_A^U, x_i \in GB_S, GB_S \cap X_p \neq \emptyset \Big\},$$
(7)

$$\underline{GB_A^U}(X_p) = \Big\{ x_i : \exists GB_S \in GB_A^U, x_i \in GB_S, GB_S \subseteq X_p \Big\}.$$
(8)

2.3. Attribute Reduction

Rough set is a powerful tool to handle fuzzy data, and we need to deal with highdimensional data through attribute reduction. By searching the minimum attribute subset which satisfies the given constraints, attribute reduction can not only reduce the dimension, but also enhance the generalization performance.

To date, various kinds of attribute reduction have been proposed for different requirements [9–13,18,32], whereas Yao et al. [33] indicated that the majority of them have analogous structures. There are two mainstream learning perspectives, i.e., supervised learning and unsupervised learning. Then, we pick the approximation quality [34] and conditional entropy [19,35–39] as two custom measures to better comprehend and investigate the essence of attribute reduction in terms of the neighborhood rough set.

2.3.1. Supervised Attribute Reduction

Supervised attribute reduction refers to the process of screening attributes using given labels in datasets so as to determine the important subsets of attributes which can best distinguish different categories.

Definition 7. *For a given decision system DS and a radius* $\delta \ge 0$, $\forall A \subseteq AT$, *the supervised approximation quality of d in terms of A is defined as*

$$\gamma_A(d) = \frac{\left| \mathbf{U}_{k=1}^q \delta_A(\mathbf{x}_k) \right|}{|\mathbf{U}|},\tag{9}$$

in which |X| is the cardinality of set X.

Apparently, it is not difficult to obtain that $\gamma_A(d) \in [0, 1]$ holds. The approximation quality reflects the proportion of samples in the lower approximation of the decision class, and it is used to describe the dependency between attributes. Note that by Definition 7, the degree of dependency increases as the value of approximation quality increases. Generally speaking, the majority of samples in *U* can be told apart from each other.

Definition 8. For a given decision system DS and a radius $\delta \ge 0$, $\forall A \subseteq AT$, the supervised conditional entropy of d based on A is defined as

$$CE_A(d) = -\frac{1}{|U|} \sum_{x \in U} |\delta_A(x) \cap [x]_d| \log \frac{|\delta_A(x) \cap [x]_d|}{|\delta_A(x)|}.$$
 (10)

It is proven that $CE_A(d) \in \left[0, \frac{|U|}{e}\right]$ holds [19]. As another important measure of the neighborhood rough set, conditional entropy reflects the discriminating performance of conditional attribute set *A* over decision attribute *d*. Following Definition 8, it is obvious that, as the value of the conditional entropy decreases, the discrimination of *A* relative to *d* increases.

Definition 9. For a given decision system DS and a constraint condition C_{ρ}^{U} , which is associated with measure ρ on the universe U, $\forall A \subseteq AT$, A is deemed as a C_{ρ}^{U} -reduct if and only if

- (1) A meets C_{ρ}^{U} ,
- (2) $\forall A_1 \subset A, A_1 \text{ does not meet } C_{\rho}^U$.

From Definition 9, it is uncomplicated to conclude that *A* is an ideal and minimal subset which satisfies the constraint condition. Without loss of generality, the constraint is closely related to the used measure. We will discuss it from the two following aspects:

- (1) If the measure is of approximation quality [28,40], the constraint condition may be $\gamma_A(d) \ge \gamma_{AT}(d)$;
- (2) If the measure is conditional entropy [41], the constraint condition may be $CE_A(d) \le CE_{AT}(d)$.

2.3.2. Unsupervised Attribute Reduction

As we all know, supervised attribute reduction depends on the labels of samples to a great extent, so it is time-consuming to obtain the labels of samples. However, unsupervised attribute reduction does not need obtain such labels.

In an unsupervised perspective, if approximate quality or conditional entropy is still needed as a measure, how to make labels for samples is an urgent problem. In order to solve the problem, Yang et al. [42] used the conditional attribute information of samples to construct pseudo-labels. Based on the pseudo-label strategy, it is not arduous to give the following definitions.

Definition 10. *For a given unsupervised decision IS and a radius* $\delta \ge 0$, $\forall A \subseteq AT$, $a \in A$, the *unsupervised approximation quality in terms of A is defined as*

$$\gamma_A = \frac{1}{|A|} \sum_{a \in A} \left(\gamma_{A-\{a\}}(d^a) \right), \tag{11}$$

in which d^a is a pseudo-label decision that records conditional attribute a to contain the pseudo-labels of samples.

In analogy with Definition 7, $\gamma_A \in [0, 1]$ apparently holds. The approximate quality in Definition 10 represents the correlation between a set of attributes and a single attribute. Naturally, the higher the value of unsupervised approximation quality is, the greater the degree of such correlation.

Definition 11. *For a given unsupervised decision IS and a radius* $\delta \ge 0$, $\forall A \subseteq AT$, $a \in A$, the unsupervised conditional entropy with respect to A is defined as

$$CE_A = \frac{1}{|A|} \sum_{a \in A} \left(CE_{A-\{a\}}(d^a) \right),$$
 (12)

in which d^{*a*} *is a derived decision that employs conditional attribute a to contain the pseudo-labels of samples.*

Similarly to Definition 8, the $CE_A \in \left[0, \frac{|U|}{e}\right]$ constant holds in an unsupervised perspective. Undoubtedly, the certainty of the pseudo-label neighborhood judgment system increases as the value of the conditional entropy decreases.

Definition 12. For a given unsupervised decision IS, a measure ρ and C_{ρ}^{U} is a constraint condition, $\forall A \subseteq AT$, A is deemed as a ρ -reduct if and only if:

- (1) A meets the constraint C_{ρ}^{U} ;
- (2) $\forall A_1 \subset A, A_1 \text{ does not meet the constraint } C_{\rho}^U$.

Analogous to Definition 9, the constraint condition determined by ρ will depend on the type of measure. The constraint condition may be $\gamma_A \ge \gamma_{AT}$ if the unsupervised approximation quality is used as a measure; it may be $CE_A \le CE_{AT}$ if the unsupervised conditional entropy is used as a measure.

3. Proposed Method

3.1. Theoretical Foundations

3.1.1. Quality-to-Entropy Ratio

Many researchers frequently use measures based on single view for attribute reduction, such as the one of supervised and unsupervised attribute reduction mentioned in Section 2.3. However, when considering only one perspective, some limitations may exist and crucial attributes cannot be effectively obtained.

Therefore, we propose a new measure which combines the supervised and unsupervised perspectives. Specifically, the supervised perspective selects the measure of approximation quality [40] and the unsupervised perspective uses conditional entropy [43]. The new measure can quantitatively describe the relationship between attributes and labels, and uncover the internal structure of data itself.

Evidently, from Definitions 7 and 11, it is known that the relationship between these two measures and the importance of attributes is completely opposite. Therefore, we adopt the form of ratio to unify the relationships. In addition, expressing the conditional entropy as an exponential function can significantly increase the relationship between conditional entropy and the importance of the attribute. The specific definition is as follows.

Definition 13. *For a given decision system DS and a radius* $\delta \ge 0$, $\forall A \subseteq AT$, *the quality-to-entropy ratio is defined as*

$$\kappa_A(d) = \frac{\gamma_A(d)}{\exp(CE_A)},\tag{13}$$

in which $\gamma_A(d)$ *is the approximation quality of d in terms of A as given in Definition 7 and* CE_A *is the unsupervised conditional entropy over A as given in Definition 11.*

According to the form of the quality-to-entropy ratio, when the value of $\gamma_A(d)$ is higher and the value of CE_A is lower, the value of $\kappa_A(d)$ is higher. From Sections 2.3.1 and 2.3.2, it can be seen that the higher the value of $\gamma_A(d)$, the greater the influence of conditional attributes on the discriminant performance of decision *d*; the lower the value of CE_A , the stronger the ability of distinguishing conditional attributes from pseudo-labels. Thus, the higher the value of $\kappa_A(d)$, the stronger the discriminant ability of conditional attributes relative to decision making. To sum up, the conclusion is in line with the ultimate goal of attribute reductions.

Theorem 1. For a given decision system DS, a radius $\delta \ge 0$, $\forall A \subseteq AT$, $\kappa_A(d) \in [0, 1]$.

Proof. $\gamma_A(d) \in [0,1]$ holds in accordance with the property of approximate quality in Definition 7. Similarly, $CE_A \in [0, \frac{|U|}{e}]$ holds in accordance with the property of conditional entropy in Definition 11, so we can infer that $\exp(CE_A) \in [1, e^{\frac{|U|}{e}}]$ holds. Immediately, the quality-to-entropy ratio $\kappa_A(d) \in [0, 1]$ holds. Specifically, if $\gamma_A(d) = 0$ and $\exp(CE_A) =$ $e^{\frac{|U|}{e}}$, $\kappa_a(d) = 0$; if $\gamma_A(d) = 1$ and $\exp(CE_A) = 1$, $\kappa_a(d) = 1$. \Box

Definition 14. *For a given decision system DS and a threshold* $\theta \in [0, 1], \forall A \subseteq AT, A$ *is defined* as a κ -reduct if and only if

- (1) $\frac{\kappa_A(d)}{\kappa_{AT}(d)} \ge \theta;$ (2) $\forall A_1 \subset A, \frac{\kappa_{A_1}(d)}{\kappa_{AT}(d)} < \theta.$

By the above definition, it is observed that, as a minimal subset of attributes, κ -reduct improves the quality-to-entropy ratio. However, the issue of how such reduct can be found urgently needs to be solved. Generally speaking, we need to evaluate the significance of attributes in AT, eliminate low-quality attributes from the reduct pool, and select qualified attributes. Based on the greedy searching for attribute reduction [44-46], Definition 15 provides a significance attribute about our proposed quality-to-entropy ratio.

Definition 15. For a given decision system DS, $\forall A \subseteq AT, \forall a \in AT - A$, we define the significance about the quality-to-entropy ratio as

$$Sig_{\kappa_a}(d) = \kappa_{A \cup \{a\}}(d) - \kappa_A(d).$$
(14)

Obviously, $Sig_{\kappa_d}(d) \in [0,1]$ holds. Definition 15 shows that the significance of conditional attributes increases as the value of $Sig_{\kappa_a}(d)$ rises. Moreover, the attributes with high significance are likely to be selected and put into the reduct pool. For example, if we assume that $Sig_{\kappa_{a_1}}(d) < Sig_{\kappa_{a_2}}(d)$, $a_1, a_2 \in AT - A$, then $\kappa_{A \cup \{a_1\}}(d) < \kappa_{A \cup \{a_2\}}(d)$ can be known. The result illustrates that we prefer a_2 to a_1 as an element in the candidate attribute subset.

3.1.2. Forward Greedy Searching to κ -Reduct Based on Granular Ball (GBFGS- κ)

Looking back on the above research, regardless of the searching methods mentioned in Section 2.3 or the quality-to-entropy ratio mentioned in Section 3.1.1, it is not difficult to obverse that: (1) searching methods obtain the information granulation based on the whole universe; and (2) in each iteration, the information granulation in the universe needs to be recalculated.

From the above reasons, the efficiency of searching methods depends on the size of the universe. Therefore, in this section, we will propose a forward greedy searching to κ -reduct based on granular ball. This strategy takes granular balls as groups to reduce the number of information granulation which need to be recalculated in each iteration. Note that Figures 1 and 2 describe its basic framework.



Figure 1. The structure of the granular ball part.



Figure 2. The structure of the κ -reduct part.

The key steps of our proposed strategy will be explained as follows:

(1) The whole universe *U* is regarded as a granular ball. Cluster each granular ball by the two-means algorithm and repeatedly cluster until all balls' purity are reached.

(2) Generate a new empty granular ball. Transfer samples into this empty granular ball if the number of samples in an existing granular ball is less than 4.

(3) Record the number of existing granular balls as n, and divide the whole universe U into n mutually asymmetric groups (U_1, U_2, \cdots, U_n) . Granular balls correspond to groups one by one.

(4) Calculate the attribute reduction result A_1 on U_1 firstly by using the quality-toentropy ratio. Then, the attribute reduction result A_2 on $U_1 \cup U_2$ will be calculated on the basis of the result A_1 .

(5) When the constraints are satisfied, a reduction result A_n is obtained which is the ultimate attribute reduction result of universe U.

Given the foregoing, it is not formidable to conclude that the forward greedy searching to the κ -reduct based on granular ball has the following benefits.

- Reduction in the number of iterations and time consumption: the merging of granular balls with fewer samples into a new ball reduces the number of iterations required, which decreases the time consumption for the subsequent calculations.
- Iterative refinement implemented in a part of the universe: based on the guiding idea, our strategy does not need to iteratively refine the information of the whole universe. This helps improve the efficiency of the attribute reduction process.
- Consideration of diversity evaluation [47,48] and complex constraints [32]: our strategy combines supervised and unsupervised learning methods to identify more significant attributes and eliminate issues that may arise from a single perspective.

According to the discussion in Section 3.1, our new strategy firstly generates granular balls by iteratively applying two-means clustering on the whole dataset. Then, we merge granular balls with fewer samples. The specific algorithm of this step is given in Algorithm 1.

Algorithm	1: Modified	generation of	of granular balls.
		()	

Input: Decision system *DS*, the threshold θ .

Output: Granular balls' index of each sample.

1 Create two granular balls: D₁ⁱ and D₂ⁱ, using the two-means clustering algorithm on the decision system DS, with *i* initialized to 1 indicating the number of iterations;

```
2 For each D_i^i do
```

- 3 The center *C* of D_i^i is the center point of all samples;
- 4 The radius *r* is the average of the distances from all samples in the D_i^i to *C*;
- 5 The purity of $pure_j^i$ can be indicative of the proportion of all samples with the same label as pellets in D_i^i ;
- 6 If $pure_i^i$ is lower than θ Then
- 7 two-means algorithm is used to decompose D_i^i ;
- 8 End
- 9 End
- 10 If the Purity of each granular ball is higher than θ Then
- 11 End;
- 12 Else
- 13 i = i + 1;
- 14 Go to **Step 2**;
- 15 End
- 16 End
- 17 Generate a new granular ball to store granular balls with less data;
- 18 Traverse each created granular ball and put samples in granular balls with less than 4 samples into the new granular ball;
- 19 Return granular balls' index of each sample.

Secondly, we select any one of granular balls, quickly identify the data in this granular ball, and obtain the attributes which meet the constraints by κ -reduct. Algorithm 2 explains the above process to us.

Algorithm 2: Forward greedy searching to *κ*-reduct (FGS-*κ*). **Input:** Decision system *DS*, the threshold θ , and a radius δ . **Output:** A κ -reduct *A*. 1 $A = \emptyset;$ ² Compute the supervised approximation quality $\gamma_{AT}(d)$; **3** For each $a \in AT$ do *a* is used to generate pseudo labels of samples in *U*; 4 5 End 6 Compute the unsupervised conditional entropy CE_{AT} ; 7 Compute quality-to-entropy ratio $\kappa_{AT}(d)$; Repeat 8 For each $a \in AT - A$ do 9 Compute $\gamma_{A \cup \{a\}}(d)$; 10 Compute $CE_{A \cup \{a\}}$; 11 Compute $\kappa_{A \cup \{a\}}(d)$; 12 End 13 Select attribute $b = argmax{Sig_{\kappa_a}(d) : \forall a \in AT - A};$ 14 15 $A = A \cup \{b\};$ Compute $\kappa_A(d)$; 16 17 Until $\frac{\kappa_A(d)}{\kappa_{AT}(d)} < heta;$ 18 Return A.

Finally, by constantly merging small universes, the reduction results obtained from previous merged small universes can be used as a guide for subsequent calculations. In other words, attributes are added on the basis of the results which obtained before merging, so as to improve the stability of reduction.

The complete algorithm of our proposed strategy is given in Algorithm 3.

Algorithm 3: Forward greedy searching to κ -reduct based on granular ball
(GBFGS- κ).
Input: Decision system <i>DS</i> , the threshold θ , and a radius δ .
Output: A reduct <i>A</i> .
1 Obtain granular balls' index of each sample by using Algorithm 1;
² Compute the number of generated granular balls, which is the value of <i>n</i> ;
³ Divide the whole universe U into n groups correspondingly based on the number
of granular balls, such as $U_1, U_2, \cdots U_n$;
4 Obtain reduct A_1 over U_1 by using Algorithm 2;
5 For $i = 2: n$ do
$A_i = A_{i-1};$
7 Repeat
8 For each $a \in AT - A_i$ do
9 Compute $\gamma_{A_i \cup \{a\}}(d)$;
10 Compute $CE_{A_i \cup \{a\}}$;
11 Compute $\kappa_{A_i \cup \{a\}}(d)$;
12 End
13 Select attribute $b = argmax{Sig_{\kappa_a}(d) : \forall a \in AT - A_i}$ with a standard;
$A_i = A_i \cup \{b\};$
15 Compute $\kappa_{A_i}(d)$;
16 Until $\frac{\kappa_{A_i}(d)}{\epsilon} < \theta$
$\kappa_{AT}(d) < 0,$
17 End
18 Return A.

In line with the process of Algorithm 3, it is not difficult to calculate the time complexity of forward greedy searching to κ -reduct based on a granular ball.

The first and foremost, in the process of generating granular balls, the universe U will be clustered into $|GB_{AT}^U|$ granular balls, resulting in a maximum of $(|U| \times |GB_{AT}^U|)$ iterations. There is one more point: when employing the κ -reduct method, pseudo-labels produced by k-means clustering have a time complexity of $O(k \cdot T \cdot |U| \cdot |AT|)$, where k is the number of clusters and T is the total number of k-means iterations. Last but not the least, in the process of recursively calculating $\kappa_A(d)$, the worst case may require adding all of the attributes in AT to the reduct pool. Obviously, the total number of times of traversing $\kappa_A(d)$ will reach $(|AT| + (|AT| - 1) + (|AT| - 2) + \cdots + 2 + 1)$. In summary, the time complexity of Algorithm 3 is $|U| \times |GB_{AT}^U| + |AT| \times |AT|^3$.

4. Experimental Analysis

4.1. Datasets

We use 16 UCI datasets for verification to demonstrate the effectiveness of our forward greedy searching to κ -reduct based on the granular ball (GBFGS- κ). Table 1 provides a thorough explanation of the various datasets.

Table 1. Dataset descriptions.

ID	Datasets	Samples	Attributes	Labels
1	Cardiotocography	2126	21	10
2	Climate Model Simulation Crashes	540	20	2
3	Diabetic Retinopathy Debrecen	1151	19	2
4	Forest Type Mapping	523	27	4
5	Ionosphere	351	34	2
6	Libras Movement	360	90	15
7	Parkinson Speech	1208	26	2
8	Pen-Based Recognition of Handwritten Digits	10,992	16	10
9	Statlog (Image Segmentation)	2310	18	7
10	Statlog (Landsat Satellite)	6435	36	7
11	Statlog (Vehicle Silhouettes)	846	18	4
12	Twonorm	7400	19	2
13	Ultrasonic Flowmeter Diagnostics-Meter D	180	43	4
14	Urban Land Cover	675	147	9
15	Wall-Following Robot Navigation	5456	24	3
16	Wisconsin Diagnostic Breast Cancer	569	30	2

4.2. Experimental Configuration

All experiments were conducted on a personal computer with Windows 10, Intel Core i7-10510U CPU(2.30 GHz) and 8.00 GB memory. The programming environment is MATLAB R2020a.

In the following experiment, the two-means algorithm was used to iteratively create granular balls, *k*-means clustering [44,49] was utilized to create pseudo-labels of samples, and quality-to-entropy ratio was the measure used in attribute reduction. It is rather remarkable that the value of *k* should be consistent with the number of decision classes in the data. In addition, the result of the neighborhood rough set largely depends on the given radius. In order to demonstrate the applicability and universality of our proposed method, all experiments employed 20 radii with a step size of 0.02, which are 0.02, $0.04, \ldots, 0.40$.

Moreover, the deduction simplification process was verified by 10-fold cross-validation. That is to say, for each radius, the samples in universe U were divided into ten groups, i.e., $U_1, U_2, \dots U_{10}$, then, nine of them were used as training groups and the rest was used as the test group. Repeat 10-fold cross-validation process for 10 times to ensure each group serves as a test group, so as to test the classification performance and obtain a reliable and stable model.

Finally, we used K-nearest neighbor (KNN, K = 3) [50,51], support vector machine (SVM) [52] and classification and regression tree (CART) [53] to compare our proposed method with six progressive algorithms in terms of attribute reduction as well as with the algorithm without applying any attribute reduction methods (No-reduct classification). The performance of the derived reducers was mainly tested from the aspects of classification stability, classification accuracy, reduced stability, and elapsed time. The attribute reduction algorithms used for comparison are as follows:

- (1) Dissimilarity Based Searching for Attribute Reduction (DBSAR) [54];
- (2) Knowledge Change Rate (KCR) [55];
- (3) Attribute Group (AG) [56];
- (4) Ensemble Selector For Attribute Reduction (ESAR) [47];
- (5) Multi-criterion Neighborhood Attribute Reduction (MNAR) [32];
- (6) Robust Attribute Reduction Based On Rough Sets (RARR) [57].

4.3. Comparison of Classification Accuracy

In this section, we will use KNN, SVM and CART to predict the test samples to weigh up the classification accuracy of each algorithm. Immediately, for attribute reduction algorithms, given a decision system *DS*, the classification accuracy applied to reduction is defined as

$$Acc_{red} = \frac{|\{x_i \in red | Pre_{red}(x_i) = d(x_i)\}|}{|U|},$$
(15)

in which $Pre_{red}(x_i)$ is the prediction label made using reduct *red* for x_i .

Table 2 displays the detailed classification accuracy results for each algorithm on 16 datasets and Figure 3 illustrates the radar charts for each dataset under the three classifiers with three different colors. The following conclusions can easily be reached by observing Table 2 and Figure 3.

- (1) For the majority of datasets, regardless of whether the KNN, SVM, or CART classifier is used, the classification accuracies related to GBSAR-κ outperform other comparison algorithms. Taking the dataset "Parkinson Speech (ID: 7)" as an example, when KNN classifier is adopted, the classification accuracies of GBSAR-κ, DBSAR, BCKCR, AG, ESAR, MMAR, RAAR, and No-reduct classification each are 0.7259, 0.7063, 0.7008, 0.7093, 0.7031, 0.7253, 0.7095, and 0.6984, respectively; when using the SVM classifier, the classification accuracies of GBSAR-κ, DBSAR, BCKCR, AG, ESAR, MMAR, RAAR, and No-reduct classification are 0.6661, 0.6532, 0.6521, 0.6548, 0.6543, 0.6639, 0.6539, and 0.6488, respectively; by employing CART, the classification accuracies of GBSAR-κ, DB-SAR, BCKCR, AG, ESAR, MMAR, RAAR, and No-reduct classification are 0.6433, 0.6429, 0.6420, 0.6413, 0.6424, 0.6307, 0.6421, and 0.6419 respectively. Therefore, the simplification derived from our GBSAR-κ can offer an effective categorization performance.
- (2) From the average classification accuracy of each algorithm, the classification accuracy associated with GBSAR- κ is comparable or even more significant than that of DBSAR, BCKCR, AG, ESAR, MMAR, RAAR, and No-reduct classification. When using the KNN classifier, GBSAR- κ 's classification accuracy is 0.8258, which is at most 32.21% higher than those of others; when the SVM classifier is utilized, GBSAR- κ 's classification accuracy is 0.8090, which is at most 27.35% higher than those of others.

Classifier	ID	GBSAR-κ	DBSAR	BCKCR	AG	ESAR	MNAR	RARR	No-Reduct Classification
	1	0.7654	0.7526	0.7576	0.7633	0.7602	0.3995	0.7532	0.7318
	2	0.9266	0.9080	0.9265	0.9120	0.9041	0.8926	0.8942	0.9059
	3	0.6147	0.6078	0.6144	0.6114	0.6103	0.5104	0.6121	0.6061
	4	0.8729	0.8670	0.8604	0.8728	0.8649	0.7086	0.8684	0.8476
	5	0.8947	0.8699	0.8790	0.8869	0.8864	0.7086	0.8944	0.8714
	6	0.6976	0.6746	0.6585	0.6674	0.6578	0.6583	0.5349	0.6544
	7	0.7259	0.7063	0.7008	0.7093	0.7031	0.7253	0.7095	0.6984
	8	0.9742	0.9670	0.9658	0.9683	0.9657	0.2732	0.2093	0.9605
KNINI	9	0.9444	0.9435	0.9418	0.9413	0.9343	0.5268	0.1589	0.9367
NININ	10	0.8897	0.8897	0.8890	0.8934	0.8890	0.5176	0.4929	0.8883
	11	0.6546	0.6464	0.6386	0.6528	0.6409	0.4722	0.6139	0.6323
	12	0.9131	0.8997	0.8958	0.8979	0.8971	0.6034	0.6065	0.8954
	13	0.8391	0.7658	0.7767	0.8014	0.7700	0.7889	0.8389	0.7681
	14	0.6801	0.6702	0.6701	0.6929	0.7004	0.5941	0.6879	0.7556
	15	0.8596	0.8619	0.8690	0.8670	0.8631	0.7314	0.5795	0.8534
	16	0.9604	0.9549	0.9511	0.9533	0.9550	0.8825	0.9602	0.9499
	Average	0.8258	0.8116	0.8122	0.8182	0.8126	0.6246	0.6509	0.8097
			↑1.75%	↑1.68%	↑0.93%	↑1.62%	↑32.21%	↑26.87%	↑1.98%
	1	0.7723	0.7460	0.7452	0.7542	0.7534	0.4579	0.7564	0.7406
	2	0.9208	0.9089	0.9200	0.9104	0.9074	0.9074	0.9077	0.9092
	3	0.6371	0.6280	0.6227	0.6285	0.6279	0.5435	0.6356	0.6210
	4	0.8717	0.8671	0.8548	0.8671	0.8650	0.6419	0.8635	0.8381
	5	0.8939	0.8369	0.8366	0.8644	0.8826	0.8000	0.8937	0.8014
	6	0.5454	0.4736	0.4717	0.4697	0.4922	0.4306	0.3679	0.4567
	7	0.6661	0.6532	0.6521	0.6548	0.6543	0.6639	0.6539	0.6488
	8	0.9353	0.9128	0.9145	0.9192	0.9186	0.3233	0.3015	0.9041
SVM	9	0.9279	0.9168	0.9149	0.9179	0.9105	0.4740	0.1498	0.9102
	10	0.8615	0.8572	0.8612	0.8611	0.8579	0.5750	0.5740	0.8466
	11	0.6342	0.6230	0.6046	0.6313	0.6141	0.3988	0.5934	0.5872
	12	0.9401	0.9266	0.9217	0.9239	0.9248	0.6624	0.6773	0.9196
	13	0.6391	0.5833	0.5617	0.5911	0.5961	0.5833	0.6389	0.6478
	14	0.7387	0.7269	0.7537	0.7613	0.7399	0.4889	0.7210	0.7267
	15	0.6808	0.6635	0.6572	0.6755	0.6602	0.6020	0.4574	0.6575
	16	0.9805	0.9688	0.9641	0.9645	0.9689	0.8754	0.9794	0.9602
	Average	0.7903	0.7683	0.7660 <u>+2</u> 179/	0.7747 10.29/	0.7734 +2.20%	0.3893	0.0337	0.7610
	1	0 8000	0 7946	0.8022	0.8046	0.8021	0.4673	0 7978	0 7906
	2	0.8090	0.7940	0.8022	0.0040	0.8021	0.4075	0.7978	0.7900
	2	0.9200	0.5071	0.6052	0.6094	0.6137	0.5783	0.6700	0.6902
	4	0.8100	0.8013	0.7943	0.8096	0.8074	0.7124	0.8014	0.8005
	5	0.8833	0.8651	0.8669	0.8664	0.8793	0.8114	0.8827	0.8613
	6	0.5038	0.4939	0.4908	0.4885	0.0798	0.4528	0.4156	0.4861
	7	0.6433	0.6429	0.6420	0.6413	0.6424	0.6307	0.6421	0.6419
	8	0.9270	0.9186	0.9244	0.9261	0.9258	0.3539	0.3045	0.9218
	9	0.9506	0.9502	0.9465	0.9501	0.9444	0.5654	0.1498	0.9416
CART	10	0.8513	0.8468	0.8493	0.8510	0.8451	0.5848	0.5910	0.8400
	11	0.6794	0.6603	0.6515	0.6748	0.6572	0.5077	0.6391	0.6641
	12	0.8205	0.8160	0.8160	0.8160	0.8168	0.5886	0.5966	0.8091
	13	0.8782	0.8467	0.8339	0.8531	0.8183	0.8056	0.8778	0.8667
	14	0.7311	0.7199	0.7567	0.7670	0.7686	0.6341	0.7583	0.7282
	15	0.9847	0.9830	0.9824	0.9844	0.9819	0.7402	0.6216	0.9817
	16	0.9286	0.9238	0.9283	0.9196	0.9250	0.8702	0.9087	0.9174
	Average	0.8090	0.7986	0.8011	0.8047	0.7996	0.6353	0.6543	0.7969
	0		↑1.31%	↑0.99%	↑0.53%	$\uparrow 1.18\%$	↑27.35%	↑23.65%	↑1.52%

 Table 2. The comparisons of the classification accuracies.

4.4. Comparison of Classification Stability

In this section, similarly to Section 4.3, we will evaluate the classification stability of each algorithm under KNN, SVM, and CART based on six advanced attribute reduction algorithms and a classification algorithm without applying any attribute reduction. Higher values of classification stability imply that the predicted label result is more stable and less susceptible to interference from the training samples.

Following the use of three classifiers on 16 datasets, Table 3 and Figure 4 show the classification stability findings of each algorithm. The following conclusions can easily be drawn in Table 3 and Figure 4.



Figure 3. Classification accuracies of three classifiers.

(1) For most datasets, our GBSAR- κ algorithm plays a leading role in classification stability compared with other algorithms. Moreover, predictions based on the features related to GBSAR- κ gain absolute advantages for some datasets. Consider the dataset "Twonorm (ID: 12)" as an example: when the KNN classifier is used, the classification

accuracies of GBSAR- κ , DBSAR, BCKCR, AG, ESAR, MMAR, RAAR, and No-reduct classification are 0.9300, 0.8934, 0.8747, 0.8744, 0.8809, 0.5300, 0.7139, and 0.8772, respectively; when adopting the SVM classifier, the classification accuracies of GBSAR- κ , DBSAR, BCKCR, AG, ESAR, MMAR, RAAR, and No-reduct classification are 0.9693, 0.9333, 0.9140, 0.9116, 0.9224, 0.5582, 0.8164, and 0.9458, respectively; when using CART, the classification accuracies of GBSAR- κ , DBSAR, BCKCR, AG, ESAR, MMAR, RAAR, and No-reduct classification accuracies of GBSAR- κ , DBSAR, BCKCR, AG, ESAR, MMAR, RAAR, and 0.9458, respectively; when using CART, the classification accuracies of GBSAR- κ , DBSAR, BCKCR, AG, ESAR, MMAR, RAAR, and No-reduct classification are 0.7723, 0.7600, 0.7531, 0.7491, 0.7564, 0.5216, 0.6803, and 0.7512, respectively. Therefore, from the standpoint of classifier stability, GBSAR- κ can indeed provide a more stable classification performance.

(2) In terms of average classification accuracy, the classification stability connected with GBSAR- κ is far superior to the other algorithms. Moreover, when employing KNN classifier, the classification stability of GBSAR- κ is 0.9015, which is at most 24.52% higher than those of other methods; the classification stability of GBSAR- κ using SVM classifier is 0.9336, which is at most 14.67% higher than those of others; the classification stability of GBSAR- κ through the use of CART classifier is 0.8301, which is at most 12.92% higher than those of others.

Table 3. The comparisons of classification stabilities.

Classifier	ID	GBSAR-ĸ	DBSAR	BCKCR	AG	ESAR	MNAR	RARR	No-Reduct Classification
	1	0.8855	0.8822	0.8750	0.8540	0.8708	0.6706	0.8762	0.8719
	2	0.9707	0.9395	0.9359	0.9308	0.9458	0.9704	0.9658	0.9380
	3	0.7943	0.7933	0.7683	0.7726	0.7814	0.4713	0.7890	0.7660
	4	0.9338	0.9295	0.9103	0.9232	0.9246	0.6990	0.9285	0.9017
	5	0.9223	0.8661	0.8633	0.8704	0.8936	0.6057	0.9363	0.8601
	6	0.8032	0.7596	0.7556	0.7538	0.7826	0.7611	0.7019	0.7517
	7	0.8165	0.8249	0.8068	0.8062	0.8142	0.8573	0.8274	0.8046
	8	0.9787	0.9677	0.9664	0.9591	0.9628	0.6905	0.8602	0.9555
KNINI	9	0.9925	0.9770	0.9738	0.9688	0.9661	0.5948	0.9909	0.9678
IN1N1N	10	0.9461	0.9427	0.9428	0.9377	0.9388	0.8824	0.8378	0.9360
	11	0.8358	0.8252	0.8156	0.8144	0.8247	0.6379	0.8350	0.8138
	12	0.9300	0.8934	0.8747	0.8744	0.8809	0.5300	0.7139	0.8772
	13	0.9453	0.8531	0.8744	0.8708	0.8997	0.7944	0.9444	0.8544
	14	0.7976	0.7657	0.7493	0.7527	0.7866	0.6519	0.8101	0.7554
	15	0.9061	0.9066	0.9023	0.8904	0.9001	0.9351	0.9043	0.9010
	16	0.9658	0.9599	0.9581	0.9530	0.9618	0.8316	0.9654	0.9542
	Average	0.9015	0.8804	0.8733	0.8708	0.8834	0.7240	0.8680	0.8642
			↑2.40%	↑3.23%	↑3.53%	↑2.05%	↑24.52%	↑3.87%	↑3.70%
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9722
	2	0.9974	0.9973	0.9845	0.9944	1.0000	1.0000	0.9994	0.9870
	3	0.9244	0.9322	0.9220	0.9207	0.9261	1.0000	0.9239	0.9160
	4	0.9669	0.9645	0.9535	0.9565	0.9638	0.7752	0.9654	0.9586
	5	0.9337	0.8710	0.8967	0.8977	0.9313	0.9143	0.9586	0.9543
	6	0.7925	0.7518	0.7608	0.7489	0.7728	0.6861	0.7393	0.7717
	7	0.9618	0.9361	0.9381	0.9259	0.9420	0.9610	0.9454	0.9396
	8	0.9708	0.9477	0.9500	0.9266	0.9458	0.6466	0.9566	0.9632
SVM	9	0.9806	0.9796	0.9748	0.9677	0.9640	0.7281	0.7455	0.9699
5 1 11	10	0.9738	0.9697	0.9732	0.9670	0.9674	0.8678	0.8876	0.9680
	11	0.8917	0.8851	0.8770	0.8747	0.8782	0.6580	0.8773	0.8568
	12	0.9693	0.9333	0.9140	0.9116	0.9224	0.5582	0.8164	0.9458
	13	0.7781	0.7575	0.7475	0.7542	0.7575	0.7556	0.7778	0.6333
	14	0.8753	0.8667	0.8574	0.8661	0.8476	0.6133	0.9049	0.8959
	15	0.9246	0.9375	0.9310	0.9014	0.9286	0.9991	1.0000	0.9434
	16	0.9968	0.9894	0.9861	0.9817	0.9857	0.8632	0.9954	0.9865
	Average	0.9336	0.9200	0.9167	0.9122	0.9208	0.8142	0.9058	0.9114
			$\uparrow 1.48\%$	$\uparrow 1.85\%$	↑2.35%	↑1.39%	$^{14.67\%}$	↑3.06%	$\uparrow 1.88\%$

Classifier	ID	GBSAR-ĸ	DBSAR	BCKCR	AG	ESAR	MNAR	RARR	No-Reduct Classification
	1	0.8366	0.8268	0.8358	0.8081	0.8282	0.7040	0.8150	0.8055
	2	0.9245	0.9013	0.9203	0.9053	0.9031	0.8852	0.8993	0.9037
	3	0.6464	0.6440	0.6401	0.6371	0.6347	0.5330	0.6394	0.6390
	4	0.9016	0.8905	0.8726	0.8879	0.8947	0.7771	0.9013	0.8776
	5	0.9144	0.8330	0.8679	0.8514	0.8939	0.9143	0.9036	0.8600
	6	0.6468	0.6396	0.6689	0.6472	0.6631	0.6556	0.6583	0.6028
CADT	7	0.6419	0.6411	0.6382	0.6338	0.6392	0.6315	0.6375	0.6259
CARI	8	0.9258	0.9087	0.9140	0.9040	0.9158	0.6673	0.9499	0.9088
	9	0.9484	0.9479	0.9422	0.9453	0.9390	0.6264	0.7732	0.9383
	10	0.8582	0.8537	0.8579	0.8565	0.8557	0.8662	0.8368	0.8457
	11	0.7363	0.7129	0.6980	0.7243	0.7275	0.6793	0.7121	0.7244
	12	0.7723	0.7600	0.7531	0.7491	0.7564	0.5216	0.6803	0.7512
	13	0.8836	0.8431	0.8369	0.8472	0.8406	0.8278	0.8833	0.8356
	14	0.7221	0.7430	0.7545	0.7552	0.7499	0.7452	0.7760	0.7628
	15	0.9811	0.9808	0.9796	0.9805	0.9791	0.9333	0.8933	0.9771
	16	0.9417	0.9312	0.9411	0.9261	0.9267	0.7947	0.9246	0.9298
	Average	0.8301	0.8161	0.8201	0.8162	0.8217	0.7352	0.8052	0.8118
	5		↑1.72%	↑1.23%	↑1.71%	↑1.02%	↑12.92%	↑3.09%	↑2.26%

Table 3. Cont.

4.5. Comparison of Reduced Stability

In this section, we will show the reduced stability of the attribute reduction corresponding to 16 datasets. The specific results are given in Table 4.

The information shown in Table 4 indicates that the reduced stability of GBSAR- κ is slightly lower than RARR, but still in a leading position. Obviously, compared with DBSAR, BCKCR, AG, ESAR, and MMAR, the average reduced stability value of GBSAR- κ is increased by 17.21%, 27.74%, 46.53%, 10.99%, and 111.21%, while it only decreases by 0.93% compared with RARR.

In general, although the reduced stability of our GBSAR- κ is not inferior to the result of RAAR for many datasets, its result is better than the six advanced algorithms in terms of attribute reduction in some cases. For instance, as far as the dataset "Climate Model Simulation Crashes (ID: 2)" is concerned, the reduced stabilities of GBSAR- κ , DBSAR, BCKCR, AG, ESAR, MMAR, and RAAR are 0.6605, 0.3265, 0.5284, 0.3483, 0.5814, 0.0545, and 0.3773, respectively. Compared with other algorithms, the result of GBSAR- κ is improved by 102.29%, 25.00%, 89.64%, 13.61%, 1111.93%, and 75.06%, respectively.

Therefore, it should be pointed out that using GBSAR- κ is more conducive to selecting attributes which are more suitable for sample changes.

4.6. Comparisons of Elapsed Time

In this section, we will compare the time taken to derive a simplification using different algorithms. The detailed results are reported in Table 5.

Following a thorough analysis of Table 5, it is not difficult to come to the findings that are listed below.

Considering the reduced stability mentioned in Section 4.5 and the reduced length conflict with each other, it can be concluded that the higher the value of reduced stability, the longer the reduced length. Apparently, the reduced length of GBSAR- κ is longer, which indicates that, in the simplification process, our algorithm needs to be strengthened in terms of the time speed.



Figure 4. Classification stabilities of three classifiers.

From the view of the average elapsed time, it is worth mentioning that the value of GBSAR- κ is 56.42% and 65.42% lower than BCKCR and RARR, respectively. Taking the dataset "Pen-Based Recognition of Handwritten Digits (ID: 8)" as an example, the speed-up ratios of GBSAR- κ algorithm reached 0.9777, 7.6960, 1.3714, 1.9355, 0.0891, and 8.4364, respectively, when the elapsed times of GBSAR- κ , DBSAR, BCKCR, AG, ESAR, MMAR, and RAAR each are 175.1957, 171.2967, 1348.3062, 240.2581, 339.0988, 15.6138, and 1478.0267 s. Therefore, the elapsed time of GBSAR- κ for attribute reduction is lower than that of AG and ESAR under some circumstances.

From the above discussion, it is observed that, even though the elapsed time of our new algorithm is better than BCKCR and RARR in some datasets, GBSAR- κ 's speed performance still has to be improved.

ID	GBSAR-ĸ	DBSAR	BCKCR	AG	ESAR	MNAR	RARR
1	0.9145	0.9280	0.8327	0.8258	0.8707	0.2000	0.9916
2	0.6605	0.3265	0.5284	0.3483	0.5814	0.0545	0.3773
3	0.9270	0.9356	0.7868	0.8032	0.8979	0.4000	0.9223
4	0.7816	0.7437	0.6017	0.6037	0.7164	0.1601	0.8677
5	0.5760	0.3647	0.2902	0.2940	0.5003	0.6308	0.9277
6	0.6587	0.2059	0.1535	0.1232	0.6033	0.1275	0.4888
7	0.8245	0.8687	0.7659	0.7281	0.8007	1.0000	0.9045
8	0.9054	0.7672	0.7155	0.5962	0.7502	0.3000	1.0000
9	0.9154	0.8968	0.8698	0.6712	0.7945	0.3000	1.0000
10	0.8609	0.7723	0.7883	0.6549	0.7442	0.6000	0.4000
11	0.8246	0.8280	0.8033	0.7261	0.8150	0.4000	0.8588
12	0.9006	0.7154	0.5958	0.5818	0.6374	0.1498	0.6000
13	0.8505	0.6466	0.6212	0.4254	0.9210	0.2051	0.9506
14	0.5577	0.3813	0.2996	0.1764	0.5346	0.4933	0.7277
15	0.8960	0.8788	0.7955	0.7183	0.8257	1.0000	1.0000
16	0.7588	0.6716	0.5824	0.4675	0.5506	0.0454	0.9157
Average	0.8008	0.6832	0.6269	0.5465	0.7215	0.3792	0.8083
		↑17.21%	↑27.74%	↑46.53%	↑10.99%	↑111 .2 1%	↓0.93%

Table 4. The reduced stabilities of deriving reducts.

Table 5. The elapsed time of deriving reducts.

ID	GBSAR- <i>k</i>	DBSAR	BCKCR	AG	ESAR	MNAR	RARR
1	19.5266	5.5508	41.9445	8.2692	10.5207	0.5752	6.3996
2	2.4537	0.2264	1.0938	0.2422	0.3454	0.5560	0.6227
3	7.0172	1.2153	7.0078	1.3858	2.0873	1.2795	4.0856
4	8.7128	0.5402	3.7415	0.6098	1.0057	0.6210	0.8102
5	2.5395	0.1988	1.0046	0.2022	0.3827	0.6178	0.9812
6	420.3743	6.1500	52.1521	7.2298	24.5777	0.8818	6.9134
7	15.1636	1.9076	11.1212	2.1158	3.4122	0.2157	5.7393
8	175.1957	171.2967	1348.3062	240.2581	339.0988	15.6138	1478.0267
9	37.3307	8.8465	61.9768	12.0828	16.6250	0.6550	26.8630
10	349.0610	169.8787	1376.8752	189.1153	322.3408	11.2199	1895.9858
11	7.5416	1.2087	8.4609	1.4569	2.1752	0.7156	1.7449
12	129.9252	33.3983	168.7067	33.4771	51.6800	6.7983	396.9944
13	21.0204	0.2190	2.1228	0.2460	0.6279	0.3357	0.8065
14	229.3661	16.9975	88.5328	19.0665	33.1863	81.6564	19.6492
15	77.3242	39.9955	296.8453	47.7270	74.3894	5.3366	528.5469
16	10.8522	0.6336	3.1715	0.6430	1.0178	2.0033	2.4340
Average	94.5878	28.6415	217.0665	35.2580	55.2170	8.0676	273.5377
		↑230.25%	↓56.42%	↑168.27%	↑71.30%	↑1072.44%	↓65.42%

5. Conclusions and Future Perspectives

In this paper, we propose a new searching strategy that differs from conventional algorithms in the following aspects. On the one hand, by automatically generating granular balls, there is no time consumption for radius optimization. On the other hand, guidance-based searching is designed to compress the attribute searching space. In addition, the quality-to-entropy ratio can overcome the limitations and predictability of the single-attribute measure method.

Through experiments on 16 UCI datasets, it is not formidable to reveal that our proposed strategy has quite a positive classification performance and strong stability in the process of exporting reduction.

Further research can be conducted for the two following aspects:

- (1) Using the fused measure may increase the time of selecting the best attribute. Therefore, more accelerators [11] can be added to further improve the efficiency and reduce the time consumption.
- (2) The searching strategy proposed in this paper is a general module. Therefore, other measures based on the rough set can be substituted for the quality-to-entropy ratio, so as to compare the classification performance under various measures.

Author Contributions: Conceptualization, J.C.; methodology, J.C.; software, M.S.; validation, J.S.; formal analysis, Y.F.; investigation, T.X.; resources, J.S.; data curation, T.X.; writing—original draft preparation, M.S.; writing—review and editing, T.X.; visualization, M.S.; supervision, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Industry-school Cooperative Education Program of the Ministry of Education (Grant No. 202101363034), the Natural Science Foundation of Jiangsu Higher Education (Grant No. 17KJB520007), and the Key Research and Development Program of Zhenjiang-Social Development (Grant No. SH2018005).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Pawlak, Z.; Skowron, A. Rudiments of rough sets. Inf. Sci. 2007, 177, 3–27. [CrossRef]
- 2. Dash, M.; Liu, H. Consistency-based search in feature selection. Artif. Intell. 2003, 151, 155–176. [CrossRef]
- 3. Dowlatshahi, M.; Derhami, V.; Nezamabadi-pour, H. Ensemble of Filter-Based Rankers to Guide an Epsilon-Greedy Swarm Optimizer for High-Dimensional Feature Subset Selection. *Information* **2017**, *8*, 152. [CrossRef]
- Chen, H.M.; Li, T.R.; Luo, C.; Horng, S.J.; Wang, G.Y. A Decision-Theoretic Rough Set Approach for Dynamic Data Mining. *IEEE Trans. Fuzzy Syst.* 2015, 23, 1958–1970. [CrossRef]
- 5. Xu, T.H.; Wang, G.Y.; Yang, J. Finding strongly connected components of simple digraphs based on granulation strategy. *Int. J. Approx. Reason.* **2020**, *118*, 64–78. [CrossRef]
- 6. Qian, Y.; Liang, J.Y.; Pedrycz, W.; Dang, C. An efficient accelerator for attribute reduction from incomplete data in rough set framework. *Pattern Recognit.* **2011**, *44*, 1658–1670. [CrossRef]
- Cheng, K.; Gao, S.; Dong, W.L.; Yang, X.B.; Wang, Q.; Yu, H.L. Boosting label weighted extreme learning machine for classifying multi-label imbalanced data. *Neurocomputing* 2020, 403, 360–370. [CrossRef]
- 8. Akram, M.; Nawaz, H.S.; Deveci, M. Attribute reduction and information granulation in Pythagorean fuzzy formal contexts. *Expert Syst. Appl.* **2023**, 222, 119794. [CrossRef]
- Jiang, Z.H.; Yang, X.B.; Yu, H.L.; Liu, D.; Wang, P.X.; Qian, Y. Accelerator for multi-granularity attribute reduction. *Knowl. Based Syst.* 2019, 177, 145–158. [CrossRef]
- 10. Chen, Y.; Yang, X.B.; Li, J.H.; Wang, P.X.; Qian, Y.H. Fusing attribute reduction accelerators. Inf. Sci. 2021, 587, 354–370. [CrossRef]
- 11. Wang, P.X.; Yang, X.B. Three-Way Clustering Method Based on Stability Theory. *IEEE Access* 2021, 9, 33944–33953. [CrossRef]
- 12. Chen, Q.; Xu, T.H.; Chen, J.J. Attribute Reduction Based on Lift and Random Sampling. Symmetry 2022, 14, 1828. [CrossRef]
- 13. Chen, Y.; Wang, P.X.; Yang, X.B.; Mi, J.S.; Liu, D. Granular ball guided selector for attribute reduction. *Knowl. Based Syst.* 2021, 229, 107326. [CrossRef]
- 14. Chen, Z.; Liu, K.Y.; Yang, X.B.; Fujita, H. Random sampling accelerator for attribute reduction. *Int. J. Approx. Reason.* 2021, 140, 75–91. [CrossRef]
- Hu, Q.H.; Pedrycz, W.; Yu, D.R.; Lang, J. Selecting Discrete and Continuous Features Based on Neighborhood Decision Error Minimization. *IEEE Trans. Syst. Man Cybern. B Cybern.* 2010, 40, 137–150. [CrossRef] [PubMed]
- 16. Liang, J.Y.; Chin, K.S.; Dang, C.Y.; Yam, R.C.M. A new method for measuring uncertainty and fuzziness in rough set theory. *Int. J. Gen. Syst.* **2002**, *31*, 331–342. [CrossRef]
- Liang, J.Y.; Shi, Z.; Li, D.Y.; Wierman, M.J. Information entropy, rough entropy and knowledge granulation in incomplete information systems. *Int. J. Gen. Syst.* 2006, 35, 641–654. [CrossRef]
- Liu, K.Y.; Yang, X.B.; Yu, H.L.; Fujita, H.; Chen, X.J.; Liu, D. Supervised information granulation strategy for attribute reduction. *Int. J. Mach. Learn. Cybern.* 2020, 11, 2149–2163. [CrossRef]
- 19. Zhang, X.; Mei, C.L.; Chen, D.G.; Li, J.H. Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy. *Pattern Recognit.* 2016, *56*, 1–15. [CrossRef]
- 20. Yao, Y.Y.; Zhang, X.Y. Class-specific attribute reducts in rough set theory. Inf. Sci. 2017, 418, 601–618. [CrossRef]
- 21. Gao, Y.; Chen, X.J.; Wang, P.X. Attribute reduction over consistent samples. CAAI Trans. Int. Syst. 2019, 14, 1170–1178. [CrossRef]

- 22. Jiang, Z.H.; Liu, K.Y.; Yang, X.B.; Yu, H.L.; Fujita, H.; Qian, Y.H. Accelerator for supervised neighborhood based attribute reduction. *Int. J. Approx. Reason.* 2020, 119, 122–150. [CrossRef]
- Zhang, W.; Miao, D.; Gao, C.; Li, F. Rough Set Attribute Reduction Algorithm for Partially Labeled Data. Comput. Sci. 2017, 44, 25–31. [CrossRef]
- Yuan, Z.; Chen, H.M.; Li, T.R.; Yu, Z.; Sang, B.B.; Luo, C. Unsupervised attribute reduction for mixed data based on fuzzy rough sets. *Inf. Sci.* 2021, 572, 67–87. [CrossRef]
- Xia, S.Y.; Liu, Y.S.; Ding, X.; Wang, G.Y.; Yu, H.; Luo, Y.G. Granular ball computing classifiers for efficient, scalable and robust learning. *Inf. Sci.* 2019, 483, 136–152. [CrossRef]
- Zhou, P.; Hu, X.G.; Li, P.P.; Wu, X.D. Online streaming feature selection using adapted Neighborhood Rough Set. Inf. Sci. 2019, 481, 258–279. [CrossRef]
- 27. Hu, Q.H.; Yu, D.R.; Xie, Z.X. Neighborhood classifiers. Expert Syst. Appl. 2008, 34, 866–876. [CrossRef]
- Liu, K.Y.; Yang, X.B.; Yu, H.L.; Mi, J.S.; Wang, P.X.; Chen, X.J. Rough set based semi-supervised feature selection via ensemble selector. *Knowl. Based Syst.* 2019, 165, 282–296. [CrossRef]
- Qian, Y.H.; Cheng, H.H.; Wang, J.T.; Liang, J.Y.; Pedrycz, W.; Dang, C.Y. Grouping granular structures in human granulation intelligence. *Inf. Sci.* 2017, 382-383, 150–169. [CrossRef]
- Yang, X.B.; Qian, Y.H.; Yang, J.Y. On Characterizing Hierarchies of Granulation Structures via Distances. *Fundam. Inform.* 2013, 123, 365–380. [CrossRef]
- Xia, S.Y.; Zhang, Z.; Li, W.H.; Wang, G.Y.; Giem, E.; Chen, Z.Z. GBNRS: A Novel Rough Set Algorithm for Fast Adaptive Attribute Reduction in Classification. *IEEE Trans. Knowl. Data Eng.* 2020, 34, 1231–1242. [CrossRef]
- Li, J.Z.; Yang, X.B.; Song, X.N.; Li, J.H.; Wang, P.X.; Yu, D.J. Neighborhood attribute reduction: A multi-criterion approach. *Int. J. Mach. Learn. Cybern.* 2019, 10, 731–742. [CrossRef]
- Yao, Y.Y.; Zhao, Y.; Wang, J.; Han, S.Q. A Model of User-Oriented Reduct Construction for Machine Learning. *Trans. Rough Sets* 2008, *8*, 332–351. [CrossRef]
- 34. Ju, H.R.; Yang, X.B.; Yu, H.L.; Li, T.J.; Yu, D.J.; Yang, J.Y. Cost-sensitive rough set approach. *Inf. Sci.* 2016, 355–356, 282–298. [CrossRef]
- Dai, J.H.; Xu, Q.; Wang, W.T.; Tian, H.W. Conditional entropy for incomplete decision systems and its application in data mining. Int. J. Gen. Syst. 2012, 41, 713–728. [CrossRef]
- Xu, J.C.; Yang, J.; Ma, Y.Y.; Qu, K.L.; Kang, Y.H. Feature selection method for color image steganalysis based on fuzzy neighborhood conditional entropy. *Appl. Intell.* 2022, 52, 9388–9405. [CrossRef]
- Sang, B.B.; Chen, H.M.; Yang, L.; Li, T.R.; Xu, W.H. Incremental Feature Selection Using a Conditional Entropy Based on Fuzzy Dominance Neighborhood Rough Sets. *IEEE Trans. Fuzzy Syst.* 2021, 30, 1683–1697. [CrossRef]
- Américo, A.; Khouzani, M.; Malacaria, P. Conditional Entropy and Data Processing: An Axiomatic Approach Based on Core-Concavity. *IEEE Trans. Inform. Theory* 2020, 66, 5537–5547. [CrossRef]
- Gao, C.; Zhou, J.; Miao, D.; Yue, X.D.; Wan, J. Granular conditional entropy-based attribute reduction for partially labeled data with proxy labels. *Inf. Sci.* 2021, 580, 111–128. [CrossRef]
- Hu, Q.H.; Liu, J.F.; Yu, D.R. Mixed feature selection based on granulation and approximation. *Knowl. Based Syst.* 2008, 21, 294–304. [CrossRef]
- 41. Mohseni, M.; Redies, C.; Gast, V. Approximate Entropy in Canonical and Non-Canonical Fiction. Entropy 2022, 24, 278. [CrossRef]
- 42. Yang, X.B.; Liang, S.C.; Yu, H.L.; Gao, S.; Qian, Y. Pseudo-label neighborhood rough set: Measures and attribute reductions. *Int. J. Approx. Reason.* **2019**, *105*, 112–129. [CrossRef]
- Li, W.Y.; Chen, H.M.; Li, T.R.; Wan, J.H.; Sang, B.B. Unsupervised feature selection via self-paced learning and low-redundant regularization. *Knowl. Based Syst.* 2021, 240, 108150. [CrossRef]
- Wang, P.X.; Shi, H.; Yang, X.B.; Mi, J.S. Three-way k-means: Integrating k-means and three-way decision. Int. J. Mach. Learn. Cybern. 2019, 10, 2767–2777. [CrossRef]
- 45. Wu, T.F.; Fan, J.C.; Wang, P.X. An Improved Three-Way Clustering Based on Ensemble Strategy. *Mathematics* 2022, 10, 1457. [CrossRef]
- 46. Ba, J.; Wang, P.X.; Yang, X.B.; Yu, H.L.; Yu, D.J. Glee: A granularity filter for feature selection. *Eng. Appl. Artif. Intell.* 2023, 122, 106080. [CrossRef]
- 47. Yang, X.B.; Yao, Y.Y. Ensemble selector for attribute reduction. Appl. Soft Comput. 2018, 70, 1–11. [CrossRef]
- 48. Chen, Y.N.; Wang, P.X.; Yang, X.B.; Yu, H.L. Bee: Towards a robust attribute reduction. *Int. J. Mach. Learn. Cybern.* 2022, 13, 3927–3962. [CrossRef]
- 49. Wang, P.X.; Yao, Y.Y. CE3: A three-way clustering method based on mathematical morphology. *Knowl. Based Syst.* **2018**, 155, 54–65. [CrossRef]
- 50. Cover, T.M.; Hart, P.E. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 1967, 13, 21–27. [CrossRef]
- 51. Fukunaga, K.; Narendra, P.M. A Branch and Bound Algorithm for Computing k-Nearest Neighbors. *IEEE. Trans. Comput.* **1975**, *C*-24, 750–753. [CrossRef]
- 52. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2011, 2, 27:1–27:27. [CrossRef]
- 53. Krzywinski, M.; Altman, N. Classification and Regression Trees. Nat. Methods 2017, 14, 755–756. [CrossRef]

- 54. Rao, X.S.; Yang, X.B.; Yang, X.; Chen, X.J.; Liu, D.; Qian, Y. Quickly calculating reduct: An attribute relationship based approach. *Knowl. Based Syst.* **2020**, 200, 106014. [CrossRef]
- 55. Jin, C.X.; Li, F.C.; Hu, Q.H. Knowledge change rate-based attribute importance measure and its performance analysis. *Knowl. Based Syst.* **2017**, *119*, 59–67. [CrossRef]
- 56. Chen, Y.; Liu, K.Y.; Song, J.J.; Fujita, H.; Yang, X.B.; Qian, Y. Attribute group for attribute reduction. *Inf. Sci.* 2020, 535, 64–80. [CrossRef]
- 57. Dong, L.J.; Chen, D.G.; Wang, N.L.; Lu, Z.H. Key energy-consumption feature selection of thermal power systems based on robust attribute reduction with rough sets. *Inf. Sci.* 2020, 532, 61–71. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.